

# The Loudest Event Statistic: General Formulation, Properties and Applications

Rahul Biswas, Patrick R. Brady, and Jolien D. E. Creighton  
*University of Wisconsin-Milwaukee, Milwaukee, WI 53201, USA*

Stephen Fairhurst  
*University of Wisconsin-Milwaukee, Milwaukee, WI 53201, USA*  
*School of Physics and Astronomy, Cardiff University, Cardiff, CF2 3YB, United Kingdom and*  
*LIGO - California Institute of Technology, Pasadena, CA 91125, USA*

The use of the loudest observed event to generate statistical statements about rate and strength has become standard in searches for gravitational waves from compact binaries and pulsars. The Bayesian formulation of the method is generalized in this paper to allow for uncertainties both in the background estimate and in the properties of the population being constrained. The method is also extended to allow rate interval construction. Finally, it is shown how to combine the results from multiple experiments and a comparison is drawn between the upper limit obtained in a single search and the upper limit obtained by combining the results of two experiments each of half the original duration. To illustrate this, we look at an example case, motivated by the search for gravitational waves from binary inspiral.

## I. INTRODUCTION

In daily life, we often estimate the birth rate, the rate of automobile fatalities, or the rate of hurricanes in the Gulf. In these cases, it is reasonably easy to determine when one event has occurred and so the best estimate is usually taken to be the number of events divided by the observation time. As physicists and astronomers, we know this is a good estimator of the rate of an underlying Poisson process. In these cases, the ability to identify events with high confidence is central to the correctness of the rate estimate.

We can carry this method over to more complicated observational situations by allowing for false positives in our identification of events. Experiments are usually designed so that the rate of real (foreground) events is higher than the rate of false positive (background) events. Hence a good estimate of the rate is obtained by counting the number of events per unit time, and making a small correction to allow for the false positives. This is the typical experimental method of estimating the rate.

In both physics and astronomy, it is common to search for very rare events in large data sets and we rely heavily on statistical methods to interpret these searches. In this paper, we discuss the problem of estimating the rate of these rare events. When real events are very rare or very weak, it is important to revisit the reasoning that underlies the standard approaches to estimating rates (and indeed other parameters). Here, we explore the effects of incorporating information about quality of observed events into the estimate of event rate. One measure of quality might be the signal to noise of the events; the louder an event, the more likely it is to be signal. Of course, more complicated measures are also possible. We simply require a rank ordering such that a larger quality implies the event is less likely to be background.

A popular method of incorporating quality information is to fix a threshold, prior to looking at the data. The threshold is often chosen to give an acceptable rate of background events in some qualitative sense. Then, the upper limit is determined by counting the number of events per unit time above

the chosen threshold and making a correction which allows for the background. Central to this method is the prescription by which the final list of events are identified.

There are many alternative criteria that might be used to determine the sample of events in an experiment. We consider using the loudest event to estimate the rate. This method was first introduced in gravitational-wave searches during the analysis of data from a prototype instrument [1]; the method was used to determine an upper limit on the rate of binary neutron star mergers in the Galaxy. Since then, the method has been used in a number of searches for gravitational waves [2, 3, 4, 5, 6, 7]. More details of this method of determining an upper limit are available in [8]. Related methods have been discussed in the context of particle physics experiments by Cousins [9] and Yellin [10].

In Sec. II, we present a general formulation of the loudest event statistic [1, 8]. We adopt the Bayesian approach which gives a posterior distribution over physical parameters based on the loudest event observed in an experiment. To provide a concrete example, in Sec. III we specialize to the case of a single unknown rate amplitude multiplying a known distribution of events. Confidence intervals based on the loudest event posterior are discussed in Sec. IV. The approach we take is sometimes called a highest posterior density interval [11]. It provides a unified approach giving an upper limit for a loudest event that is due to noise with high probability and a confidence interval (bounded away from zero) when the loudest event is foreground with high probability. In real experiments there are many systematic uncertainties; we discuss marginalization over uncertainties in Sec. V. Finally, we explain how to combine the results from multiple experiments using the loudest event method and show that the resulting upper limit is independent of the order of the experiments. This discussion also leads naturally to an investigation of the effect on an upper limit if a single search is split into two parts. In Sec. VII, we make some general comments on the results obtained in this paper. In Appendix A, we consider the application of a Feldman-Cousins unified approach to obtaining a frequentist upper limit using the loudest observed events. A comparison between the loudest event method and an event counting

method of obtaining an upper limit is given in Appendix B for a toy problem.

Notation: We use the following notation throughout this paper. The loudest event statistic variable is denoted  $x$ . The experimentally measured values of a quantity are denoted by a circumflex accent; e.g., the experimentally measured value of the loudest event statistic is given by  $\hat{x}$ . Probability distributions and other quantities related to an experimental background appear with a subscripted 0; e.g.,  $P_0(x)$  is the background probability distribution for the loudest event statistic. The symbol  $B$  is used to describe the collective information about the experimental background in conditional probability distributions that are contingent on this information; thus  $P(x|B)$  is the distribution of the loudest event statistic in an experiment that includes a background.

## II. GENERAL FORMULATION OF LOUDEST EVENT STATISTIC

Consider a search of experimental data for a rare Poisson process. The output of this search is a set of candidate events which have survived all cuts applied during the analysis. At first, suppose that all the events are *foreground* events. Assume that these events can be ranked according to a single parameter  $x$ , such as a signal-to-noise ratio, in such a way that the probability that the search will detect an event increases with increasing  $x$ . For simplicity we will call this parameter the *loudness* parameter and we will say that the event with the largest value of  $x$  is the *loudest event*. If the mean number of events expected during the course of the experiment with the ranking statistic value above  $x$  is given by  $\nu(x)$ , then the probability of observing no events above a given value of  $x$  is

$$P(x) = e^{-\nu(x)}. \quad (1)$$

However, if the experiment can produce *background* events, then the probability that there are no events, either foreground or background, louder than  $x$  is

$$P(x|B) = P_0(x)e^{-\nu(x)} \quad (2)$$

where we have used  $B$  to indicate that the probability depends on the background and the factor  $P_0(x)$  is the probability of obtaining zero background events louder than  $x$ .

The mean number of events expected during the course of the experiment,  $\nu(x)$ , depends on the duration of the experiment, the rate of events, and the ability of the experiment to detect events that occur. The sensitivity of the search is encoded in the efficiency which is the probability that an event will have a loudness value greater than or equal to  $x$ . The efficiency depends on  $x$  as well as a set of parameters, collectively denoted by  $\theta$ , that determine the detectability of a source. For example,  $\theta$  may include such things as the sky position, orientation, distance, etc., of an astrophysical source. We write the efficiency as  $\varepsilon(x, t, \theta)$ . (The sensitivity of the experiment may change with time; hence the explicit dependence on  $t$  in the efficiency.) The rate of events depends on the parameters  $\theta$  that describe the source population as well as on physical parameters, collectively denoted by  $\mu$ , that we are interested in measuring or constraining by means of the experiment. We write the rate of events as  $R(\theta, \mu)$ . With this factorization, the mean number of events expected can be expressed as

$$\nu(x, \mu) = \int_0^T dt \int d\theta \varepsilon(x, t, \theta) R(\theta, \mu), \quad (3)$$

where  $T$  is the total observation time.

We can substitute our expression for the rate (3) into Eq. (2) to obtain the probability that there are zero events in the data with a loudness statistic value greater than  $x$  as

$$P(x|\mu, B) = P_0(x) \exp \left\{ - \int_0^T dt \int d\theta \varepsilon(x, t, \theta) R(\theta, \mu) \right\}. \quad (4)$$

Furthermore, the probability of the loudest event occurring between  $x$  and  $x + dx$  is given by  $p(x|\mu, B) dx$  where

$$\begin{aligned} p(x|\mu, B) &= \frac{d}{dx} P(x|\mu, B) \\ &= p_0(x) \left[ 1 - \left( \frac{P_0(x)}{p_0(x)} \right) \int_0^T dt \int d\theta \frac{d\varepsilon(x, t, \theta)}{dx} R(\theta, \mu) \right] e^{-\nu(x, \mu)} \end{aligned} \quad (5)$$

and  $p_0(x) = dP(x)/dx$ . Notice that the probability distribution contains two factors: an exponential decay that is determined by the (foreground) rate of events and a shape factor comprising two terms.

After performing an experiment, we are interested in obtaining a distribution for the model parameters that govern the

rate. To do this, we calculate a Bayesian posterior distribution for these parameters,  $\mu$ , given the observations. This distribution is denoted  $p(\mu|\hat{x}, B)$ , where  $\hat{x}$  is the value of the observed loudest event, and it is derived using Bayes' law:

$$p(\mu|\hat{x}, B) = \frac{p(\mu) p(\hat{x}|\mu, B)}{\int d\mu p(\mu) p(\hat{x}|\mu, B)} \quad (6)$$

where  $p(\mu)$  is the prior probability distribution on the model parameters. In many circumstances, the parameters  $\mu$  may be further divided into a set of particular interest  $\mu_I$  and others of less interest  $\mu_{II}$ . By integrating Eq. (6) over the unwanted parameters  $\mu_{II}$ , one obtains the posterior distribution

$$p(\mu_I|\hat{x}, B) = \frac{\int d\mu_{II} p(\mu_I, \mu_{II}) p(\hat{x}|\mu_I, \mu_{II}, B)}{\int d\mu p(\mu) p(\hat{x}|\mu, B)}. \quad (7)$$

In Sec. V we consider this procedure of marginalization over unwanted, or nuisance, parameters in more detail.

To bound the parameters of interest at a given confidence level  $\alpha$ , one integrates Eq. (7) over some region  $\Omega(\mu_I)$  such that

$$\alpha = \int_{\Omega(\mu_I)} d\mu p(\mu_I|\hat{x}, B). \quad (8)$$

In general, the difficult part is selecting the region  $\Omega(\mu_I)$ , especially in more than one dimension. There are several ways to do this: for example, one could marginalize over all but one of the parameters thus reducing the problem to a one-dimensional integral; or select the smallest volume  $\Omega(\mu_I)$  that gives the required probability. This is sometimes called a highest posterior density interval [11]. In Sec. IV, we investigate the properties of this type of rate interval based on the loudest event method.

### III. UPPER LIMIT ON UNKNOWN RATE AMPLITUDE

We have obtained the general expression for the posterior probability distribution of the parameters  $\mu$  governing an astrophysical model based on an observed loudest event. In practice, the details of obtaining either a rate upper limit or a confidence interval on the model parameters will depend upon the details of the astrophysical model and its dependence upon the variables  $\mu$ . In this section, we simplify to the situation where the rate is dependent upon a single parameter  $\mu$ , an overall unknown Poisson mean number of events, so that

$$R(\theta, \mu) = \frac{\mu f(\theta)}{T} \quad (9)$$

where  $T$  is the observation time and  $f(\theta)$  is the distribution of events as a function of  $\theta$ .

We can use this form of the rate to simplify the general expression for the posterior. To begin, we introduce the quantity

$$\epsilon(x) = \frac{1}{T} \int_0^T dt \int d\theta \epsilon(x, t, \theta) f(\theta) \quad (10)$$

which can be regarded as an averaged detection efficiency: the probability that a foreground event will have a loudness parameter greater than  $x$ . Then, the mean number of events with ranking statistic above  $x$  is  $\nu(x) = \mu \epsilon(x)$ , and (at least in principle)  $\epsilon(x)$  is known. The posterior distribution is determined by substituting Eqs. (9) into (3) and using Eqs. (5), (6), and (10) to obtain

$$p(\mu|\hat{\epsilon}, \hat{\Lambda}) = \frac{p(\mu) p_0(\hat{x}) (1 + \mu \hat{\epsilon} \hat{\Lambda}) e^{-\mu \hat{\epsilon}}}{\int d\mu p(\mu) p_0(\hat{x}) (1 + \mu \hat{\epsilon} \hat{\Lambda}) e^{-\mu \hat{\epsilon}}} \quad (11)$$

where the function  $\Lambda(x)$  is given by

$$\Lambda(x) = \left( \frac{-1}{\epsilon(x)} \frac{d\epsilon(x)}{dx} \right) \left( \frac{p_0(x)}{P_0(x)} \right)^{-1} \quad (12)$$

and a hat over a function indicates evaluation at  $\hat{x}$ . The quantity  $\Lambda(x)$  is a measure of the relative probability of detecting a single event with loudness parameter  $x$  versus such an event occurring due to the experimental background; in particular  $\hat{\Lambda} \rightarrow 0$  in the limit that the loudest event is definitely from the background and  $\hat{\Lambda} \rightarrow \infty$  in the limit that the loudest event is definitely from the foreground. Note that if the possibility that the loudest event could be from a background is ignored, the posterior distribution,  $p(\mu|\hat{\epsilon}) \propto \mu \hat{\epsilon} e^{-\mu \hat{\epsilon}}$  is peaked away from zero and vanishes as  $\mu \rightarrow 0$ ; that is, the posterior distribution will be inconsistent with zero foreground events.

For the rest of the paper we take a uniform prior except in Section VI in which we consider using the posterior from a first experiment as a prior for a second experiment. It should be noted that only power law priors on the rate, including the uniform prior, do not introduce a timescale into the problem. Power laws with powers greater than -1 are needed to avoid a required low-rate cutoff, which again would introduce a natural timescale.

Let us evaluate the upper limit making use of a uniform prior,

$$p(\mu) = \text{const.} \quad (13)$$

While this distribution is not normalizable, we can introduce a cutoff at large  $\mu$  (well above the expected number of events during the given experiment) in order to render it normalizable. Physically, this is a reasonable choice of prior if there is no information available about the expected value of  $\mu$ . Furthermore, the posterior distribution is insensitive to the value of the cutoff provided it is sufficiently large. For the uniform prior, the posterior distribution in Eq. (11) evaluates to

$$p(\mu|\hat{\epsilon}, \hat{\Lambda}) = \frac{\hat{\epsilon}}{1 + \hat{\Lambda}} (1 + \mu \hat{\epsilon} \hat{\Lambda}) e^{-\mu \hat{\epsilon}}. \quad (14)$$

It is straightforward to show that the distribution in Eq. (11) will be peaked away from zero if and only if  $\hat{\Lambda} > 1$ ; the mode of the distribution is

$$\mu_{\text{peak}} = \begin{cases} 0 & \hat{\Lambda} \leq 1 \\ (\hat{\Lambda} - 1)/\hat{\Lambda} \hat{\epsilon} & \hat{\Lambda} > 1. \end{cases} \quad (15)$$

If  $\hat{\Lambda} > 1$  then one might take this as an indication of a non-zero rate. The extent to which this is true is explored in Sec. IV.

We integrate Eq. (14) to obtain an upper limit at confidence level  $\alpha$  by solving

$$\begin{aligned} \alpha &= \int_0^\mu d\mu' p(\mu'|\hat{\epsilon}, \hat{\Lambda}) \\ &= 1 - \left[ 1 + \frac{\mu \hat{\epsilon} \hat{\Lambda}}{1 + \hat{\Lambda}} \right] e^{-\mu \hat{\epsilon}(\hat{x})} \end{aligned} \quad (16)$$

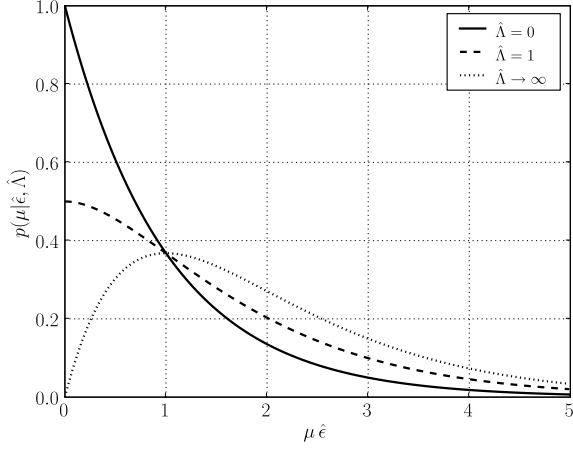


FIG. 1: The posterior probability density function  $p(\mu|\hat{\epsilon}, \hat{\Lambda})$  on the Poisson mean  $\mu$ , assuming a uniform prior, and a fixed value  $\hat{\epsilon}$  of the efficiency evaluated at the loudest event  $\hat{x}$ . The three curves correspond to three different values of  $\hat{\Lambda}$ : a)  $\hat{\Lambda} = 0$  (solid line), the loudest event is definitely background and the distribution is exponential; b)  $\hat{\Lambda} = 1$  (dashed line), the transitional case where the distribution peaks at zero but the derivative vanishes there; c)  $\hat{\Lambda} \rightarrow \infty$  (dotted line), the loudest event is definitely from the foreground, the distribution is peaked away from zero.

for  $\mu$ . It has been shown in [8] that setting the background to zero yields a conservative rate limit. In the Bayesian analysis, however, this yields a posterior probability distribution function which is peaked away from zero, and goes to zero at zero rate. This is clearly seen in Fig. 1 which shows the posterior distribution for three values of  $\hat{\Lambda}$  including  $\hat{\Lambda} \rightarrow \infty$ . This is not surprising as we have neglected the background, in which case the existence of a loudest event implies a non-zero rate. Although this does not invalidate the upper limit, it does mean that the posterior would not serve as a suitable prior for a future experiment, as it is inconsistent with a zero rate. Nevertheless, it is still possible to obtain the upper limit as

$$\frac{\mu_{90\%}}{T} = \frac{3.890}{T\epsilon(\hat{x})}. \quad (17)$$

Similarly, the no-foreground limit can be obtained by taking  $\hat{\Lambda} = 0$ . In this case, the 90% confidence limit tends to

$$\frac{\mu_{90\%}}{T} = \frac{2.303}{T\epsilon(\hat{x})}. \quad (18)$$

Finally, we can consider the transitional case  $\hat{\Lambda} = 1$ :

$$\frac{\mu_{90\%}}{T} = \frac{3.272}{T\epsilon(\hat{x})}. \quad (19)$$

The posterior distribution for the Poisson mean  $\mu$  for these three possibilities is shown in Fig. 1.

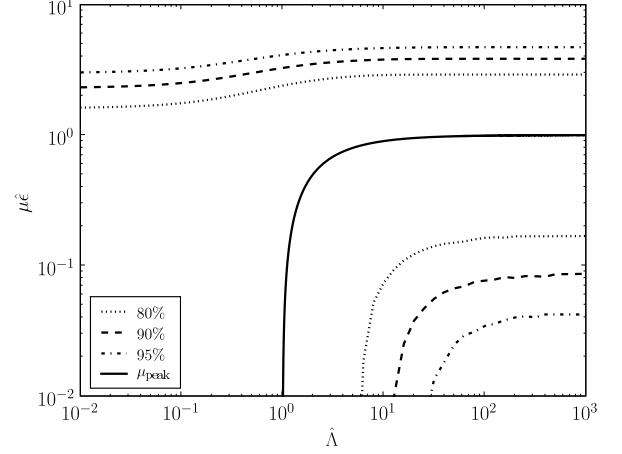


FIG. 2: The graph shows the behavior of the lower and upper boundaries of the interval,  $\mu_1$  and  $\mu_2$  respectively, as a function of  $\hat{\Lambda}$ . They are plotted for three different values of the confidence level  $\alpha$  of 80%, 90% and 95%. The peak  $\mu_{\text{peak}}$  (solid line) approaches zero as  $\hat{\Lambda}$  approaches zero. As  $\hat{\Lambda} \rightarrow 0$ ,  $\mu_2$  agrees with the no foreground upper limit treated above.

#### IV. CONFIDENCE INTERVAL ON UNKNOWN RATE AMPLITUDE

In Sec. III, we derived the upper limit on the Poisson mean  $\mu$  based on the loudest event. However, in the case where the value of  $\hat{\Lambda}$  is large (likely to be foreground), one might prefer to obtain a rate interval rather than an upper limit. For a uniform prior, the mode  $\mu_{\text{peak}}$  of the posterior distribution for the Poisson mean, given in Eq. (15), is non-zero whenever  $\hat{\Lambda} > 1$ . Furthermore, in this case,  $\mu_{\text{peak}}$  asymptotes to  $1/\hat{\epsilon}$  for large values of  $\hat{\Lambda}$  as one might expect. How significant an indicator of a non-zero rate is having the peak of rate distribution be non-zero? In order to examine this idea more precisely, we describe a method of constructing a rate interval using the loudest event statistic which provides a unified approach similar to Feldman and Cousins [12].

At some confidence level  $\alpha$ , an interval is given by  $[\mu_1, \mu_2]$  such that

$$\int_{\mu_1}^{\mu_2} p(\mu|\hat{\epsilon}, \hat{\Lambda}) d\mu = \alpha. \quad (20)$$

A supplementary condition is required to select a unique interval: we identify the interval which minimizes  $|\mu_2 - \mu_1|$  and contains the mode of the distribution (or zero for  $\hat{\Lambda} < 1$ ). This condition clearly results in  $\mu_1 = 0$  for small values of  $\hat{\Lambda}$ , i.e. when the loudest event was likely to have arisen from the background, the rate interval on the process we wish to constrain includes zero rate.

For the uniform prior, the dependence of  $\mu_1$ ,  $\mu_2$  and  $\mu_{\text{peak}}$  on  $\hat{\Lambda}$  are shown in Fig. 2. For  $\hat{\Lambda} < 1$ ,  $\mu_{\text{peak}} = 0$  and consequently  $\mu_1 = 0$ , as expected. However, for a significant range of  $\hat{\Lambda} > 1$ , even though the rate distribution is peaked away



from zero,  $\mu_1 = 0$  indicating that (at the given confidence) the rate interval still includes zero.

We can determine the precise value of  $\hat{\Lambda}$  at which  $\mu_1$  becomes non-zero. For fixed  $\hat{\Lambda}$  and  $\hat{\epsilon}$ , Eq. (20) gives  $\mu_2$  implicitly as a function of  $\mu_1$ . The minimal interval condition is then just

$$\frac{d[\mu_2(\mu_1) - \mu_1]}{d\mu_1} = 0. \quad (21)$$

Substituting  $\mu_1 = 0$  into Eqs. (20) and (21), we obtain two equations which depend on  $\mu_2$  and  $\hat{\Lambda}$ . As an example, consider a 90% confidence interval. In this case,  $\mu_1$  becomes non-zero, and the interval is bounded away from the origin, at value of  $\hat{\Lambda} \simeq 11.56$ . This corresponds to  $\mu_2 \simeq 3.807/\epsilon(\hat{x})$ . This result is in good agreement with the values obtained numerically in Fig. 2.

It is interesting to note that the 90% confidence interval still includes zero for a wide range of  $\hat{\Lambda}$  that give posterior distributions peaked away from zero. Figure 3 provides a concrete example of the posterior when  $\hat{\Lambda} = 10$ ; the 90% confidence interval still includes zero.

## V. MARGINALIZATION OVER UNCERTAINTIES

The expected mean number of detected events,  $\nu(x, \mu)$  in Eq. (3), is dependent upon the frequency of events and their amplitude distribution as well as the sensitivity of the search which is performed. In many cases, neither of these quantities will be precisely known. For example, the efficiency of an experiment is often measured via Monte-Carlo methods and therefore suffers from uncertainties due to the finite number of trials. If we expand our understanding of the parameters  $\mu$  to further parametrize the uncertainties that can arise in the underlying models and in measurements of efficiency, it is natural to marginalize over these uncertainties before computing an upper limit or rate interval. Just as the marginalization over uninteresting physical parameters [given in Eq. (7)] requires a prior distribution to be specified, the same is true of the uncertainties. This prior distribution would typically reflect the systematic and statistical errors estimate for the experiment.

### A. Marginalization over uncertainties in $\epsilon$

As a particular example, consider the problem of the unknown rate amplitude presented in Sec. III and assume there is some uncertainty associated with the value of  $\hat{\epsilon} = \epsilon(\hat{x})$ . Typically, one might choose the prior to be a normal distribution of the variate  $\epsilon$  peaked around the estimate value of  $\hat{\epsilon}$ . It is, however, unphysical for the rate to be zero, so the distribution would need to be truncated. A more natural choice is a log-normal distribution, for which the logarithm of  $\epsilon$  would be normally distributed, thereby guaranteeing that  $\epsilon$  is positive.

Here, we choose to make use of the  $\gamma$ -distribution, primarily because it can be analytically integrated. The  $\gamma$ -distribution is similar in shape (for small standard deviation) to both the Gaussian and log-normal distributions and in addition takes only non-negative values. The  $\gamma$ -distribution is given by

$$p(\epsilon; k, \theta) = \frac{\epsilon^{(k-1)} e^{-\epsilon/\theta}}{\theta^k \Gamma(k)} \quad (22)$$

where  $\Gamma(k)$  is the Gamma function. The mean is  $\bar{\epsilon} = k\theta$  while the standard deviation is  $\sigma_\epsilon = k^{1/2}\theta$ . Therefore, the fractional standard deviation,  $\sigma_\epsilon/\bar{\epsilon} = k^{-1/2}$  tends to zero in the limit as  $k \rightarrow \infty$ , whereby we expect to recover the unmarginalized results. Note that the  $\gamma$ -distribution is a distribution over the domain  $\epsilon \in [0, \infty)$  while the efficiency actually takes values only between 0 and 1. If  $k$  is large then the  $\gamma$ -distribution is sharply peaked about its mean value and we can ignore this issue.

The marginalized distribution is calculated by integrating over  $\epsilon$ ,

$$p(\mu|k, \hat{\epsilon}, \hat{\Lambda}) = \left[ \int_0^\infty d\epsilon p(\epsilon; k, \theta) p(\mu|\epsilon, \hat{\Lambda}) \right]_{\theta=\hat{\epsilon}/k}. \quad (23)$$

where we set the value of the parameter  $\theta$  so that the mean of the  $\gamma$ -distribution equals the observed efficiency  $\hat{\epsilon}$ , and where  $k$  is a measure of the fractional uncertainty in the value of  $\hat{\epsilon}$ . Making use of the distribution (11) for the Poisson mean parameter and the expression for the  $\gamma$ -distribution given above, we obtain the marginalized distribution for integer values of  $k$

$$p(\mu|k, \hat{\epsilon}, \hat{\Lambda}) = \frac{\hat{\epsilon}}{(1 + \hat{\Lambda})} \left[ \frac{1}{(1 + \mu\hat{\epsilon}/k)^{k+1}} + \frac{\mu\hat{\epsilon}\hat{\Lambda}(1 + 1/k)}{(1 + \mu\hat{\epsilon}/k)^{k+2}} \right]. \quad (24)$$

In the limit that  $k \rightarrow \infty$ , we recover the previous distribution for  $\mu$  as expected.

In order to examine the effect of marginalization, in Fig. 3 we plot the unmarginalized posterior distribution for  $\hat{\Lambda} = 10$  along with three distributions obtained by marginalizing over different size systematic errors or uncertainties. These distri-

butions are obtained from (24) with values of  $k = 100, 16$  and 4 corresponding to errors of 10%, 25% and 50% respectively. As the systematic error increases, the posterior distribution for the Poisson mean parameter gets broader; the value of the probability density function increases for large values of the Poisson mean parameter. This causes an increase in the upper limit. Without taking into account any uncertainties, the 90%

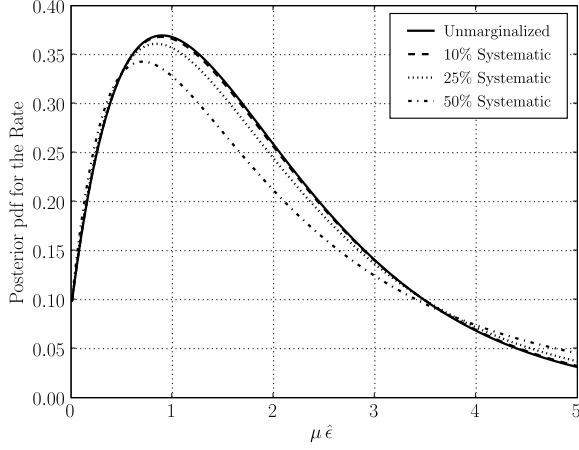


FIG. 3: The posterior probability density function on the Poisson mean parameter  $\mu$  for different sizes of systematic error, where  $\hat{\epsilon}$  is the efficiency evaluated at the loudest event  $\hat{x}$ . The curves were generated assuming a uniform prior and using  $\hat{\Lambda} = 10$ . The solid line corresponds to the unmarginalized probability density function. The dot-dashed line gives the distribution marginalized over a 10% systematic uncertainty (equivalently  $k = 100$  for the  $\gamma$ -distribution). With this level of uncertainty, the marginalized distribution is barely changed from the original. The dotted and dashed lines show the posterior for 25% ( $k = 16$ ) and 50% ( $k = 4$ ) systematic errors. As the systematic error increases the distribution broadens and consequently the upper limit increases.

confidence upper limit is  $3.796/\epsilon(\hat{x})$ . For 10% systematic error, this increases only slightly to  $3.850/\epsilon(\hat{x})$  while for 25% and 50% this increases further to  $4.147/\epsilon(\hat{x})$  and  $5.434/\epsilon(\hat{x})$  respectively. In Figure 4 we plot the upper limit as a function of the systematic error for four different values of  $\Lambda$ . The results are qualitatively similar to what was seen before — marginalizing over uncertainties will increase the upper limit and the larger the errors, the larger the effect.

### B. Marginalization over uncertainties in $\Lambda$

In many cases, there will also be uncertainties in the precise value of  $\hat{\Lambda} = \Lambda(\hat{x})$ . These can be marginalized over in the same way as described above. Since the  $\hat{\Lambda}$  dependence of the distribution (11) is straightforward, this can be done explicitly. For concreteness, let us take a uniform prior, in which case the posterior distribution is given by Eq. (14). Then, given a probability distribution  $p(\Lambda)$ , the marginalized distribution is

$$p(\mu|\hat{\epsilon}) = \int d\Lambda p(\Lambda) p(\mu|\hat{\epsilon}, \Lambda) \quad (25)$$

In this case, the above integral is straightforward. Specifically, let us define

$$\xi = \int d\Lambda p(\Lambda) \frac{\Lambda}{(1 + \Lambda)}. \quad (26)$$

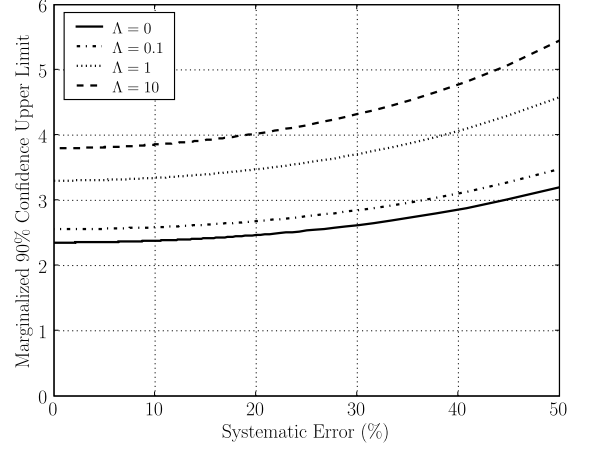


FIG. 4: The 90% confidence upper limit versus the size of the systematic error which is marginalized over (equivalent to  $k^{-1/2}$  in the  $\gamma$ -distribution discussed in the text). The limit is plotted for four different values of  $\hat{\Lambda}$ : 0, 0.1, 1, 10. In all cases, the upper limit increases with larger systematic error.

Then, the posterior distribution following marginalization over  $\Lambda$  is given by

$$p(\mu|\hat{\epsilon}) = \hat{\epsilon} [(1 - \xi) + \mu\hat{\epsilon}\xi] e^{-\mu\hat{\epsilon}} \quad (27)$$

where  $\xi$  contains all of the dependence of the posterior on the marginalized background.

Suppose that  $\Lambda$  is distributed with expectation value  $\hat{\Lambda}$  and variance  $\sigma_{\Lambda}^2$ . Then, to leading order,

$$\xi \approx \frac{\hat{\Lambda}}{1 + \hat{\Lambda}} - \frac{\sigma_{\Lambda}^2}{(1 + \hat{\Lambda})^3}. \quad (28)$$

From this, we notice two things. First, even if the fractional uncertainties in  $\hat{\Lambda}$  are of order unity, when  $\hat{\Lambda} \gg 1$  or  $\hat{\Lambda} \ll 1$ , the second term is small compared to the first and can be ignored. Second, marginalizing over  $\Lambda$  only serves to decrease the value of  $\xi$  relative to the unmarginalized case. This is equivalent to reducing the likelihood that the loudest event is foreground and consequently will reduce the upper limit. Therefore, it is possible to neglect the marginalization of  $\Lambda$  as this is a conservative thing to do.

## VI. COMBINING RESULTS FROM MULTIPLE EXPERIMENTS

When performing a series of experiments, there is a very natural way to combine the results in a Bayesian manner. As discussed above, the calculation of a Bayesian upper limit requires the specification of a prior probability distribution for the rate  $\mu$ . When a previous experiment has been performed, it is natural to use the posterior from the first experiment as the prior for the second. It is straightforward to show that the results are independent of the order of the experiments. (This

does not depend upon the loudest event, rather it is a general Bayesian result.) Begin by recalling that

$$p(\mu|\hat{x}_1) = \frac{p(\mu) p(\hat{x}_1|\mu)}{\int d\mu p(\mu) p(\hat{x}_1|\mu)}. \quad (29)$$

For the second search, simply use  $p(\mu|\hat{x}_1)$  as the prior to obtain the posterior distribution on  $\mu$  given the observations in both the first and second experiments:

$$p(\mu|\hat{x}_1, \hat{x}_2) = \frac{p(\mu) p(\hat{x}_1|\mu) p(\hat{x}_2|\mu)}{\int d\mu p(\mu) p(\hat{x}_1|\mu) p(\hat{x}_2|\mu)}. \quad (30)$$

This is clearly symmetric in  $\hat{x}_1$  and  $\hat{x}_2$ . It is straightforward to see that marginalization over nuisance parameters (see Sec. V) preserves this symmetry.

Let us consider this in more detail. If the first search was performed using a uniform prior, the posterior is given by Eq. (11) with a loudest event value  $\hat{x}_1$  observed in the first search. Furthermore, in the event that the loudest event is most likely background, one expects  $\hat{\Lambda}_1 \ll 1$ . Then, we can conservatively rewrite the posterior as

$$p_{\text{conservative}}(\mu|\hat{\epsilon}_1, \hat{\Lambda}_1) = \hat{\epsilon}_1 \hat{\Lambda}_1 e^{-\mu \hat{\epsilon}_1 (1 - \hat{\Lambda}_1)} \quad (31)$$

where  $\hat{\epsilon}_1 = \epsilon_1(\hat{x}_1)$ ,  $\hat{\Lambda}_1 = \Lambda_1(\hat{x}_1)$ , and we have made use of the fact that

$$1 + \mu \hat{\epsilon}_1 \hat{\Lambda}_1 \leq e^{\mu \hat{\epsilon}_1 \hat{\Lambda}_1}. \quad (32)$$

It is straightforward to show that the rate limit at a given confidence level  $\alpha$  inferred using this posterior is necessarily larger than that obtained using the original distribution. In this sense, the alternative distribution is conservative and the distribution has been cast as an exponential.

Therefore, in the second search, it is natural to use an exponential prior,

$$p(\mu) = \kappa e^{-\kappa \mu}. \quad (33)$$

To obtain the posterior distribution obtained when the exponential prior is used, it is beneficial to re-define  $\Lambda(x)$  as

$$\Lambda_\kappa(x) = \left( \frac{-1}{\epsilon_\kappa(x)} \frac{d\epsilon_\kappa(x)}{dx} \right) \left( \frac{p_0(x)}{P_0(x)} \right)^{-1} \quad (34)$$

where

$$\epsilon_\kappa(x) = \epsilon(x) + \kappa \quad (35)$$

includes the exponential scale constant from the prior distribution. Then, the posterior distribution is given by

$$p(\mu|\hat{\epsilon}_\kappa, \hat{\Lambda}_\kappa) \propto (1 + \mu \hat{\epsilon}_\kappa \hat{\Lambda}_\kappa) e^{-\mu \hat{\epsilon}_\kappa} \quad (36)$$

with  $\hat{\epsilon}_\kappa = \epsilon_\kappa(\hat{x})$  and  $\hat{\Lambda}_\kappa = \Lambda_\kappa(\hat{x})$ . As before, the posterior distribution is peaked away from zero if  $\hat{\Lambda}_\kappa > 1$ . In addition, the distribution is identical to that obtained using a uniform prior, only now the search efficiency is effectively  $\epsilon(\hat{x}) + \kappa$ .

### A. Splitting a search

Next, let us consider the effect of taking a single search and splitting it into two halves, which can be combined to produce an upper limit in the manner described above. Naively, it appears that splitting the search will give a lower rate limit, since we will be using a quieter loudest event for half the search. If this were the case, then it would seem that splitting the search into ever shorter searches would lower the upper limit indefinitely. As we shall see, the result is not so clear cut, and it depends critically upon the foreground and background distributions  $\epsilon(x)$  and  $P_0(x)$ .

Consider an experiment performed for some given time  $T$ , and assume that both the foreground and background rates are constant over time. We would then like to compare the (expected) upper limit from the full search to that obtained by splitting the data in two parts of length  $T_1$  and  $T_2$  and calculating a combined upper limit from the two searches. Let us assume, without loss of generality, that the loudest event overall in the search occurs in the first half of the search with a statistic value of  $\hat{x}_1$ , and the loudest event in the second half of the search has a statistic value  $\hat{x}_2$ . Then, we can calculate the upper limit from the search (taking it as a single entity) and from the split search.

The posterior for the single search is given by

$$p(\mu|\hat{x}_1, B) = \frac{p(\mu) [1 + \mu \epsilon(\hat{x}_1) \Lambda(\hat{x}_1)] \exp\{-\mu \epsilon(\hat{x}_1)\}}{\int d\mu p(\mu) [1 + \mu \epsilon(\hat{x}_1) \Lambda(\hat{x}_1)] \exp\{-\mu \epsilon(\hat{x}_1)\}} \quad (37)$$

while for the split search, the likelihood for each part is proportional to

$$p(\hat{x}_i|\mu, B) \propto [1 + \mu \eta_i \hat{\epsilon}_i \hat{\Lambda}_i] e^{-\mu \eta_i \hat{\epsilon}_i} \quad (38)$$

where  $i = 1, 2$  label the two parts of the search,  $\hat{\epsilon}_i = \epsilon_i(\hat{x}_i)$ ,  $\hat{\Lambda}_i = \Lambda_i(\hat{x}_i)$  and  $\eta_i = T_i/T$  is the fraction of the total observation time that is contained in the each interval;  $\eta_1 + \eta_2 = 1$ . Then the combined posterior distribution for the split search is

$$p(\mu|\hat{x}_1, \hat{x}_2, B) = \frac{p(\mu) \left[1 + \mu\eta_1\hat{\epsilon}_1\hat{\Lambda}_1\right] \left[1 + \mu\eta_2\hat{\epsilon}_2\hat{\Lambda}_2\right] \exp\{-\mu[\eta_1\hat{\epsilon}_1 + \eta_2\hat{\epsilon}_2]\}}{\int d\mu p(\mu) \left[1 + \mu\eta_1\hat{\epsilon}_1\hat{\Lambda}_1\right] \left[1 + \mu\eta_2\hat{\epsilon}_2\hat{\Lambda}_2\right] \exp\{-\mu[\eta_1\hat{\epsilon}_1 + \eta_2\hat{\epsilon}_2]\}}. \quad (39)$$

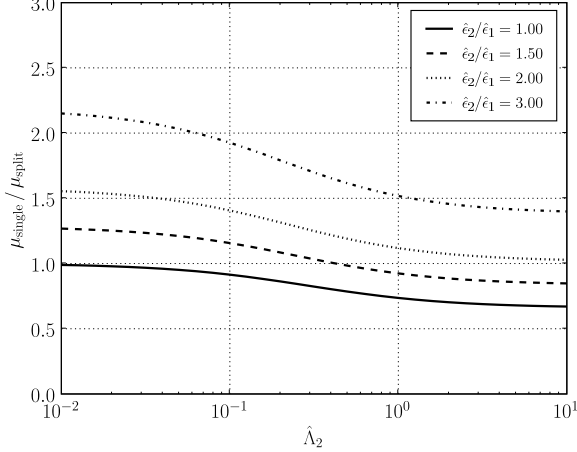


FIG. 5: The ratio of  $\mu_{\text{single}}$  to  $\mu_{\text{split}}$  as a function of  $\hat{\Lambda}_2$  for several values of  $\hat{\epsilon}_2/\hat{\epsilon}_1$ . The data presented uses a uniform prior on  $\mu$ ,  $\eta_i = \frac{1}{2}$ ,  $\epsilon(x) = \epsilon_i(x)$  and  $\Lambda_1(x) = \Lambda_2(x)$ . The figure was generated for  $\Lambda(\hat{x}_1) = 0.5\Lambda_1(\hat{x}_1) = 1$ . In general, there is only a weak dependence on this value; the curves steepen a little for smaller value of  $\Lambda(\hat{x}_1)$ , but look qualitatively similar. Note also that for most sensible choices of amplitude statistic  $x$ , one expects  $\hat{\Lambda}_2 \leq \hat{\Lambda}_1$ . The plot is extended to  $\hat{\Lambda}_2 = 10$  for completeness.

We can now compare the posterior distributions for the single and split search. In general, the efficiencies  $\epsilon_i(x)$  and the likelihoods  $\Lambda_i(x)$  could be different for the two parts. Nevertheless, it follows directly from Eq. (10) that  $\epsilon(x) = \eta_1\epsilon_1(x) + \eta_2\epsilon_2(x)$ . Therefore, in the split search, the exponential decay term is at least as large as for the single search, with equality only if  $x_2 = x_1$ . This tends to make the upper limit obtained in the split search smaller than that of the single search. In contrast, the polynomial prefactor is always more significant for the split search (i.e. it grows more steeply with  $\mu$ ). This tends to make the upper limit larger. So splitting the search will lead to a larger limit when  $\hat{\epsilon}_2 = \hat{\epsilon}_1$  and  $\hat{\Lambda}_2 = \hat{\Lambda}_1$ . Meanwhile if  $\hat{\epsilon}_2 \gg \hat{\epsilon}_1$  and  $\hat{\Lambda}_2 \ll \hat{\Lambda}_1$ , the split search will give a numerically smaller limit.

Assuming a uniform prior on  $\mu$ , we compare the 90%-confidence upper limits on the Poisson mean obtained from a single search with the limit obtained from a split search with  $\eta_1 = \eta_2 = 1/2$ . The results are shown in Fig. 5 for several choices of  $\hat{\epsilon}_2/\hat{\epsilon}_1$  under the assumptions that  $\epsilon(x) = \epsilon_i(x)$  and  $\Lambda_1(x) = \Lambda_2(x)$ . While not the most general case, these assumptions are reasonable in the context of an experiment with the same apparatus and background noise sources. In the limit  $\hat{\Lambda}_2 \rightarrow 0$ , we find  $\mu_{\text{single}} > \mu_{\text{split}}$  as expected. As  $\hat{\Lambda}_2$  increases, the second event is less likely to be background and

so the rate limit from the split search can become bigger than that obtained in the single search. While this makes intuitive sense, the result depends on the particular observed outcomes of the experiment.

When  $\hat{\Lambda}_1 \ll 1$  and  $\hat{\Lambda}_2 \ll 1$ , the posterior distribution for the single search can be approximated conservatively as

$$p(\mu|\hat{x}_1, B) \simeq \epsilon(\hat{x}_1)[1 - \Lambda(\hat{x}_1)]e^{-\mu\epsilon(\hat{x}_1)[1 - \Lambda(\hat{x}_1)]} \quad (40)$$

while the posterior for the split search becomes

$$p(\mu|\hat{x}_1, \hat{x}_2, B) \simeq c(\hat{x}_1, \hat{x}_2)e^{-\mu c(\hat{x}_1, \hat{x}_2)} \quad (41)$$

where

$$c(x_1, x_2) = \hat{\epsilon}_1\eta_1(1 - \hat{\Lambda}_1) + \hat{\epsilon}_2\eta_2(1 - \hat{\Lambda}_2). \quad (42)$$

If we further assume that both the foreground and background are Poisson distributed over the entire search, then  $\epsilon_i(x) = \epsilon(x)$  and  $\eta_i\Lambda_i(x) = \Lambda(x)$ . Within the context of these assumptions, it is then easy to write down the upper limit for each distribution. In particular,

$$\mu_{\text{single}} = \frac{-\ln(1 - \alpha)}{\epsilon(\hat{x}_1)[1 - \Lambda(\hat{x}_1)]} \quad (43)$$

for the single search; for the split search

$$\mu_{\text{split}} = \frac{-\ln(1 - \alpha)}{c(\hat{x}_1, \hat{x}_2)}. \quad (44)$$

Hence, the single search will give a smaller upper limit if

$$\Lambda_2(\hat{x}_2) > \left[1 - \frac{\epsilon(\hat{x}_1)}{\epsilon(\hat{x}_2)}\right]. \quad (45)$$

Once again, the comparison between the single and split search is sensitive to the precise nature of the foreground, background and observed results.

## VII. DISCUSSION

The loudest event statistic is just one method of taking account of the quality of an event in the interpretation of a search. In this paper, we have presented further exploration of the method including the discussion of marginalization over uncertainties in the input model. The Bayesian approach allows simple accounting of these uncertainties by integrating them out.

We also showed how the method could be used to determine a rate interval. Once again, this is not the most powerful method of determining an interval (in the sense that using



more than one event would lead to a more strongly peaked distribution and, consequently, a narrower interval). Nevertheless, the approach shows that a rate interval arises when the likelihood that the event is signal becomes large enough.

Finally, we presented a discussion of combining the results from multiple searches to determine a single upper limit. It was shown that the limit obtained by combining two searches of equal duration is, in general, different to the limit obtained by performing a single search of equivalent duration. What conclusion to draw from this is unclear since the notion of better depends on the true value of the rate being explored.

Even though physicists have a deep appreciation for probabilistic phenomena in nature, it is often tempting to talk about better upper limits by using one method or another. This is, of course, a flawed approach. In fact, it is the experiment that one should choose not the statistical method. Nevertheless, some experiments may be more powerful than others. For example, it would be ill-conceived to use the loudest event method to determine a rate interval in an experiment which is likely (in the sense of prior probability) to generate more than one loud event that could be considered to arise from the phenomenon of interest. Indeed, these considerations lead back to an experiment more like the standard threshold approach.

### Acknowledgments

We would like to acknowledge many useful discussions with members of the LIGO Scientific Collaboration inspiral analysis group which were critical in the formulation of the methods and results described in this paper. This work has been supported in part by NSF grants PHY-0200852 and PHY-0701817; PRB is grateful to the Research Corporation for support by a Cottrell Scholar Award; SF was funded in part by the Royal Society. LIGO was constructed by the California Institute of Technology and Massachusetts Institute of Technology with funding from the National Science Foundation and operates under cooperative agreement PHY-0107417. This paper has LIGO Document Number LIGO-P070076-00-Z.

### APPENDIX A: FREQUENTIST APPROACH TO THE LOUDEST EVENT STATISTIC

This paper has explored the loudest event statistic from the Bayesian point of view. In this appendix we consider an alternative approach: the construction of frequentist confidence intervals using the loudest event statistic. The construction of frequentist upper limits is almost trivial (see [8]). More interesting is the application of the method of Feldman and Cousins [12] for a unified approach to constructing confidence intervals. We restrict attention here to the case in which only the unknown rate amplitude  $\mu$  is to be bounded.

We briefly summarize the Neyman approach to constructing confidence belts: For each fixed value of  $\mu$ , an interval  $[x_1, x_2]$  is constructed such that the probability of observing a loudest event  $x$  in this interval is equal to the desired confi-

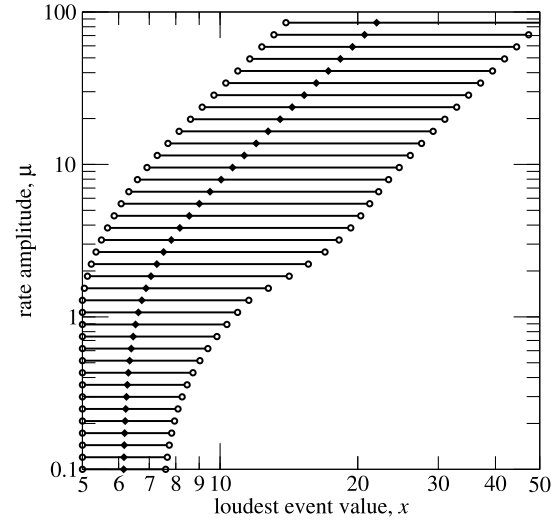


FIG. 6: The  $\alpha = 90\%$  confidence belt constructed using the procedure of Feldman and Cousins for the example described in the text. For each value of  $\mu$ , the probability of obtaining a loudest event value of  $x$  in the interval given by the solid line bounded by open circles is 90%. The solid diamond shows the point in this interval where  $R(x)$  is the greatest. For an observed loudest event value  $\hat{x}$ , the interval on  $\mu$  is the intersection of the vertical line  $x = \hat{x}$  with this belt. Note that for small values of  $\hat{x}$  an upper limit on  $\mu$  is obtained while for larger values of  $\hat{x}$  the interval on  $\mu$  is bounded away from zero. The division between an upper limit and an interval that excludes zero occurs at a value of  $x_{90\%}$  for which  $P_0(x_{90\%}) = 90\%$ .

dence level  $\alpha$ :

$$P(x_2|\mu) - P(x_1|\mu) = \alpha \quad (\text{A1})$$

where

$$P(x|\mu) = P_0(x)e^{-\mu\epsilon(x)}. \quad (\text{A2})$$

The collection of such intervals then defines a confidence belt; for any observed value of the loudest event  $\hat{x}$ , the belt covers a range of values of  $\mu$ ,  $[\mu_1, \mu_2]$ , which is the desired interval on the rate amplitude parameter. In order to construct the confidence belt, a supplementary condition is needed in order to uniquely define the interval. For example, to obtain a confidence belt that always yields upper limits, choose  $x_2 = \infty$ . Then the belt is defined as the interval  $[x_1, \infty)$  where  $x_1$  satisfies

$$1 - P(x_1|\mu) = \alpha \quad (\text{A3})$$

for each value of  $\mu$ . Then, given an observed loudest event value  $\hat{x}$ , the rate amplitude interval is  $[0, \mu_2]$  where

$$1 - P(\hat{x}|\mu_2) = \alpha \quad (\text{A4})$$

or

$$\mu_2 = -\frac{\ln(1 - \alpha) - \ln P_0(\hat{x})}{\epsilon(\hat{x})}. \quad (\text{A5})$$

As mentioned in [8], this procedure has the pathology that if  $P_0(\hat{x}) < 1 - \alpha$  then the interval on  $\mu$  is empty.

The unified approach of Feldman and Cousins provides a different supplementary condition for defining the interval  $[x_1, x_2]$  for fixed  $\mu$ . In the Feldman and Cousins approach, the interval is constructed using a function  $R(x)$  so that  $R(x) \geq R_{\min}$  for  $x \in [x_1, x_2]$  and  $R(x) < R_{\min}$  for  $x$  outside the interval where  $R_{\min} = \min\{R(x_1), R(x_2)\}$ . The function  $R(x)$  is chosen to be the likelihood ratio

$$R(x) = \frac{p(x|\mu)}{p(x|\mu_{\text{peak}})} \quad (\text{A6})$$

$$= \begin{cases} [1 + \mu\epsilon(x)\Lambda(x)]e^{-\mu\epsilon(x)} & \Lambda(x) \leq 1 \\ \frac{[1 + \mu\epsilon(x)\Lambda(x)]e^{-\mu\epsilon(x)}}{\Lambda(x)e^{1/\Lambda(x)-1}} & \Lambda(x) > 1 \end{cases} \quad (\text{A7})$$

where

$$p(x|\mu) = \frac{dP(x|\mu)}{dx} = p_0(x)[1 + \mu\epsilon(x)\Lambda(x)]e^{-\mu\epsilon(x)} \quad (\text{A8})$$

and  $\mu_{\text{peak}}$  is given by Eq. (15).

As an illustration, we compute the confidence belt according to the Feldman and Cousins procedure for the following example: We take  $\epsilon(x) = (x_{\min}/x)^3$  and  $P_0(x) = 1 - \exp(x_{\min} - x)$  with  $x_{\min} = 5$ ; only values  $x > x_{\min}$  are realizable. The  $\alpha = 90\%$  confidence belt is shown in Fig. 6. Notice that there is a well-defined interval on  $\mu$  for any observed loudest event value  $\hat{x}$  possible, that is, the Feldman and Cousins procedure produces confidence belts that are free of the pathology described above when upper limit confidence belts are constructed. Furthermore, the interval on  $\mu$  is an upper limit for small loudest event values  $\hat{x} < x_{90\%}$ , but becomes an interval which excludes zero for  $\hat{x} > x_{90\%}$  where  $P_0(x_{90\%}) = 90\%$ .

It is interesting to consider the behavior of the confidence intervals for large values of  $x$ . In this regime,  $\Lambda(x) \gg 1$ ,  $P_0(x) \simeq 1$  so  $R(x) \simeq \mu\epsilon(x)e^{-\mu\epsilon(x)+1}$  and  $P(x|\mu) \simeq e^{-\mu\epsilon(x)}$ . The confidence belt at fixed  $\mu$  is given by the interval  $[x_1, x_2]$  that satisfies  $R(x_1) = R(x_2)$  and  $P(x_2|\mu) - P(x_1|\mu) = \alpha$ . Therefore,  $x_1$  and  $x_2$  satisfy the coupled equations  $\mu\epsilon(x_1)e^{-\mu\epsilon(x_1)} = \mu\epsilon(x_2)e^{-\mu\epsilon(x_2)}$  and  $e^{-\mu\epsilon(x_2)} - e^{-\mu\epsilon(x_1)} = \alpha$ . For  $\alpha = 90\%$ ,  $\mu\epsilon(x_1) = 3.932$  and  $\mu\epsilon(x_2) = 0.08381$ . Consequently, the 90% confidence rate amplitude interval  $[\mu_1, \mu_2]$  for large  $\hat{x}$  will be given by  $\mu_1 = 0.08381/\epsilon(\hat{x})$  and  $\mu_2 = 3.932/\epsilon(\hat{x})$ . Notice the ratio  $\mu_2/\mu_1 = 46.91$  is fixed: the fractional uncertainty in the value of  $\mu$  does not improve as  $\hat{x}$  increases. This demonstrates that for experiments in which events in the low-background region are expected, the loudest event statistic will not give strong constraints on the event rate. As emphasized in the introduction, the loudest event statistic is best suited to problems in which the anticipated event rate is very low and foreground and background events are expected to have comparable amplitudes.

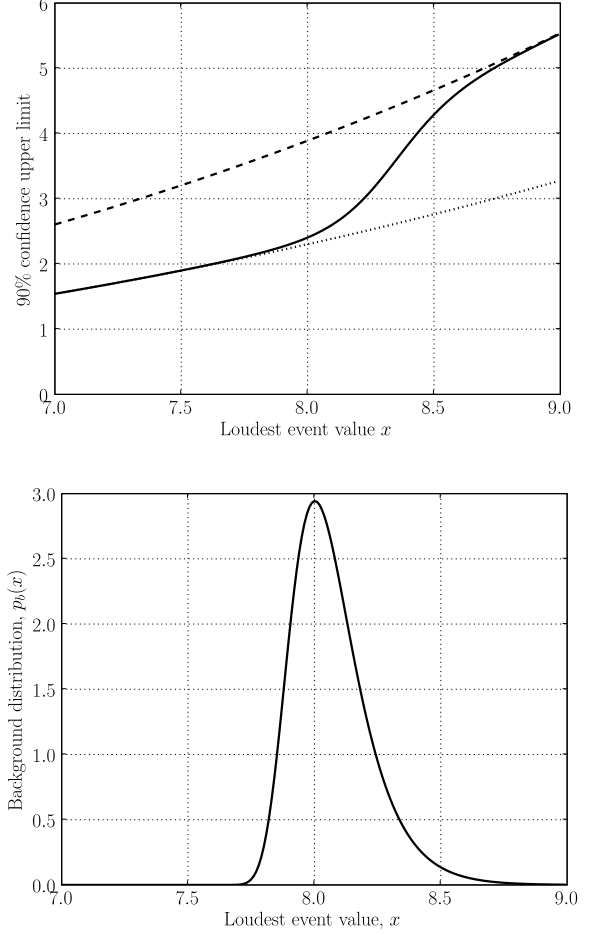


FIG. 7: a) The upper limit as a function of the observed loudest event. The solid line shows the value of the upper limit as a function of  $\hat{x}$ . The dotted and dashed lines are given by  $2.303/\epsilon(\hat{x})$  and  $3.890/\epsilon(\hat{x})$ . We see that the upper limit transitions smoothly from one to the other. At low values of  $\hat{x}$ , the loudest event is very much consistent with the background,  $\hat{\Lambda} \approx 0$  and the upper limit is close to the dotted line. For larger values of  $\hat{x}$  the loudest event is more consistent with foreground,  $\hat{\Lambda} \rightarrow \infty$ , and the upper limit is more consistent with the dashed line. b) The probability distribution for the loudest event assuming that it is drawn from the background distribution,  $p_0(x)$ . Multiplying the upper limit curve by this distribution and integrating over  $x$  gives the expected value of the upper limit if the loudest event is from the background.

## APPENDIX B: COMPARISON WITH FIXED THRESHOLDS

Let us compare the loudest event statistic against a fixed threshold approach. The loudest event prescription can be applied to any form of background, provided the required quantities in Eq. (12) can be measured or estimated. In many experiments, one might expect the background events above a statistic value  $x$  to be Poisson distributed, with mean  $\nu_0(x)$  where  $\nu_0$  is a non-increasing function of  $x$ . Then, it follows

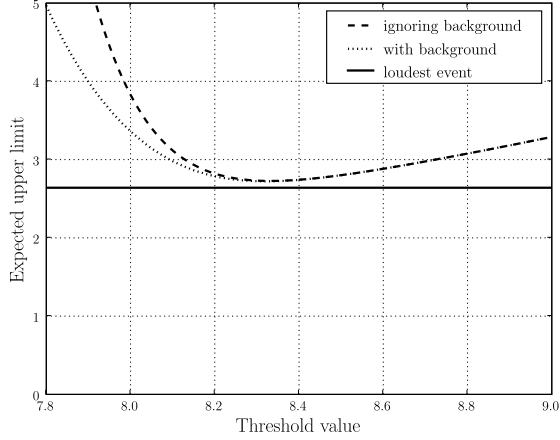


FIG. 8: Figure showing the expected upper limit as a function of the fixed threshold. The dashed line shows the upper limit obtained when ignoring the background, while the dotted line includes the background contribution. For large values of the threshold where the expected background is small, both limits approach  $2.303/\epsilon(x^*)$  as expected. For low values of  $x^*$ , there is a good chance of many events above threshold which leads to a worse upper limit. The balance occurs at around a threshold value of  $x^* = 8.3$ . For reference, we also plot a horizontal line showing the expected upper limit from the loudest event. Interestingly, the loudest event will, on average, outperform the fixed threshold for any value of the threshold.

directly that

$$\begin{aligned} P_0(x) &= e^{-\nu_0(x)} \\ p_0(x) &= \left| \frac{d\nu_0(x)}{dx} \right| e^{-\nu_0(x)} \\ p_0(x)/P_0(x) &= \left| \frac{d\nu_0(x)}{dx} \right|. \end{aligned} \quad (\text{B1})$$

We work with an example where the mean  $\nu_0(x) = e^{(8^2 - x^2)/2}$  and the foreground is distributed as  $\epsilon(x) = (8/x)^3$ . These choices are natural in the context of a search for gravitational waves from coalescing binaries where the background is  $\chi^2$  distributed with two degrees of freedom, while the foreground is uniformly distributed in volume, and the signal

strength (and hence statistic value  $x$ ) are inversely proportional to the distance [13]. The normalizations of these functions are chosen for simplicity so that  $\nu_0(8) = \epsilon(8) = 1$ . The main feature of these distributions, however, is simply that they are both decreasing functions of  $x$ , and that the background decreases more rapidly than the foreground. The value of the upper limit as a function of the actual loudest event is shown in Fig. 7a. The upper limit transitions smoothly from the zero foreground limit (at low values of  $x$ ) to zero background limit (at large values of  $x$ ). Figure 7b shows the distribution  $p_0(x)$ . This corresponds to the expected distribution of for the loudest event if it is due to the background. Then, by multiplying the upper limit by the expected distribution for the loudest event and integrating, we obtain the expected upper limit. In this example it is 2.64.

For comparison, the upper limit for a fixed threshold is presented in Figure 8. When calculating the upper limit for a fixed threshold, one simply counts the number of events  $\hat{n}$  above the chosen threshold  $x^*$  and obtains a limit

$$\frac{\mu_{90\%}}{T} = \frac{F(\hat{n})}{\epsilon(x^*)T} \quad (\text{B2})$$

where  $F(n)$  is a known function for each integer  $n$  (see, for example, [14] for more details). In particular, when zero events are observed above the threshold,  $F(0) = 2.303$ . When performing a fixed threshold search, it is possible to take into account the expected background and, much as for the loudest event, neglecting to do so will lead to a conservative result. In Fig. 8, we show the expected upper limit as a function of the threshold.

Clearly, in this example, the loudest event statistic is preferable to a fixed threshold, as it will provide a better expected upper limit value than the fixed threshold for *any* value of the threshold (with or without the background). We note that this result is specific to the details of the example under consideration; the key feature is that the background rate is a very steep function of  $x$ . Indeed, in [8], the same example was considered, but with an expected background of unity at  $\hat{x} = 4.5$  rather than  $\hat{x} = 8$ , leading to a small range of values where the fixed threshold does beat the loudest event. However, as emphasized in that paper the attraction of the loudest event is that it is unnecessary to fix a threshold ahead of performing the search — the search itself determines the threshold.

[1] B. Allen et al., Phys. Rev. Lett. **83**, 1498 (1999).  
[2] B. Abbott et al. (LIGO Scientific Collaboration), Phys. Rev. D **69**, 122001 (2004), gr-qc/0308069.  
[3] B. Abbott et al. (LIGO Scientific Collaboration), Phys. Rev. D **72**, 082002 (2005), gr-qc/0505042.  
[4] B. Abbott et al. (LIGO Scientific Collaboration), Phys. Rev. D **72**, 082001 (2005), gr-qc/0505041.  
[5] B. Abbott et al. (LIGO Scientific Collaboration), Phys. Rev. D **73**, 062001 (2006), gr-qc/0509129.  
[6] B. Abbott et al. (LIGO Scientific), Phys. Rev. **D69**, 082004 (2004), gr-qc/0308050.  
[7] X. Siemens et al., Phys. Rev. **D73**, 105001 (2006), gr-

qc/0603115.  
[8] P. R. Brady, J. D. E. Creighton, and A. G. Wiseman, Class. Quant. Grav. **21**, S1775 (2004).  
[9] R. Cousins, Nucl. Instrum. Meth. **A337**, 557 (1994).  
[10] S. Yellin, Phys. Rev. **D66**, 032005 (2002), physics/0203002.  
[11] W.-M. Yao et al., J. Phys. G: Nucl. Part. Phys. **33**, 1 (2006).  
[12] G. J. Feldman and R. D. Cousins, Phys. Rev. D **57**, 3873 (1998), physics/9711021.  
[13] P. Brady and S. Fairhurst, gr-qc (2007), arXiv:0707.2410.  
[14] W. M. Yao et al. (Particle Data Group), J. Phys. **G33**, 1 (2006).